THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 57, No. 4, April 1978
Printed in U.S.A.

Loop Plant Modeling:

Statistical Analyses of Costs in Loop Plant Operations

By D. M. DUNN and J. M. LANDWEHR

(Manuscript received October 10, 1977)

The Serving Area Concept (SAC) *involves a new procedure for the design and administration of the loop plant to reduce operating costs. Two major problems facing a loop plant engineer considering conversion to* SAC *are determining which areas should be converted (and in what order) and assessing the savings resulting from the conversion. This paper presents methodology and data analysis results useful for solving such problems. The data analyzed are from the Prototype District and measure a large number of facility related problems both before and after conversion to* SAC*. A cost penalty measure, based on observed facility problems, is calculated for a given area using data collected in that area over a certain period of time. The before conversion data are characterized and modeled in order to quantify the uncertainty, in the form of a confidence interval, associated with this cost penalty. Confidence intervals are useful to decide appropriate sizes for the data collection areas, appropriate lengths of time for data collection, as well as for comparing the results between two or more areas. The effect of conversion to* SAC *on the cost penalty measure is also examined. It is found that after conversion costs are much lower than before conversion costs, but that costs continue to decrease for at least 9 to 12 months after conversion takes place. The analysis and results presented here yield methods and guidelines to be used for data collection and analysis in other districts. These can help in reliably choosing areas for conversion to* SAC *which will maximize savings.*

I. INTRODUCTION AND SUMMARY

Investment decisions in the loop plant, like most such investment decisions in the Bell System, are dependent on careful analyses and the data which underlie these analyses. This paper describes detailed studies of a large body of data measuring several kinds of loop plant operations and costs. The cost measures used are based on the *Facility Analysis Plan for Outside Plant* (FAP); this plan, described and discussed in Ref. 1, gives methods for managing the loop plant. The results of this paper contain guidelines for the use of certain FAP measures, as well as insights into related characteristics of the data.

The data analyzed here are from the Prototype District Project,[2] a major effort undertaken to analyze those operating costs of a district that can be controlled by changes in the design or administration of the loop network. This involved a nearly three year study of the Passaic District of New Jersey Bell Telephone Company. Passaic is an urban area with some small business, scattered apartments, and large old houses. Many sections were converting from single- to multiple-family dwellings. Much of the existing loop plant was congested and had maintenance problems. Thus, conversion to the Serving Area Concept (SAC)[3] was considered appropriate for much of the district. This conversion involves departures from dedicated plant design and multipled plant design.[4] Serving area interfaces, which are basically large boxes containing cable pair interconnect points, are installed in appropriate places in the network. Then cable pairs are permanently connected from the interface to the customer, and complements of feeder pairs from the central office to the interface are supplied as needed. The Facility Analysis Plan, developed from the Prototype District Project, gives methods for determining when and where conversion to SAC is appropriate.

The Prototype District Data Base[5] is the key to tracking district activities. Each month over 50,000 measurements of district operations involving facility related problems were recorded. (Many of these measurements were zero.) Data were retained by 50-pair complement by month for that part of the district undergoing extensive conversion to serving areas. Data are available from April 1973 through December 1975.

There are many procedures in the Facility Analysis Plan to aid in understanding costs and potential savings in the management of loop plant. Among the concepts involved are allocation areas,[1,4] which are geographical regions used for tracking operating costs and cable usage. Allocation areas are also basic units of plant for planning additions or changes in the network such as conversion to SAC. Therefore, in order to trigger the need for treatment of the network these areas are initially ranked on the basis of facility problems in each area. This ranking is based on a weighted linear combination of facility problems normalized

by the number of assigned pairs in the area. The weights are costs associated with the individual problem items and together yield a "Cost Penalty Per Assigned Pair" (CPPAP). In Ref. 1, the Normalized Yearly Marginal Operating Cost, which is a generalization of CPPAP, is used as a basis for their discussion. Other cost calculations include the "Plant Stabilization Analysis Form" and the "CUCRIT" analysis to compute the rate of return associated with a given investment strategy. While these other cost calculations are important and relevant to FAP, the focus of this paper is on the CPPAP calculation and its component parts.

Three specific reasons motivate the choice of CPPAP for analysis here. First, it is the initial form used to analyze data in FAP and as such holds an important position. Second, the cost calculations for CPPAP are linear combinations of observed quantities and hence directly interpretable. Third, CPPAP does not require any special factors (e.g., "improvement factor") as are needed in most of the other measures.

The general purpose of this paper is to give insight into facets of these data relating to the conversion of selected allocation areas to SAC which took place during the Prototype District Project. Two important problems to the loop plant engineer are to determine which of the allocation areas should be converted (and in what order), and to assess the savings resulting from the conversion. The data analysis addresses these problems by modeling the variability of the FAP data. The uncertainty associated with projected savings is found to decrease as the serving areas become larger (in assigned pairs) and the data collection period increases.

An exploratory analysis of the before, during, and after conversion cost measure and its components (Section II) shows that the cost measure varies widely both across areas and time. Assignment changes, cable troubles, and defective pairs contribute the most to the level and variability. A detailed statistical analysis of the before conversion cost data in Section III is used as a basis to develop confidence intervals (Section IV) on the "true" cost penalty. These intervals quantify the uncertainty associated with an observed cost penalty for a given area. They are useful to decide appropriate sizes for the data collection areas, appropriate lengths of time for data collection, as well as for comparing the results between two or more areas. Moreover, confidence intervals show the trade-off between the size of the data collection area and the data collection period.

Finally, the effect of the conversion on the cost measure is examined in Section V. A regression equation is developed which models the after conversion costs in terms of before and during conversion variables as well as the time since conversion. The major result shows that costs continue to decrease after conversion takes place. In order to get an adequate measure of the savings associated with conversion to SAC, one must collect data for at least nine to twelve months after conversion.

It should be noted (before proceeding with the data analysis) that much of the work described was also performed on other savings measures including the rate of return. The same techniques which are shown for CPPAP were found useful, but for brevity their results are not shown.

II. GENERAL CHARACTERISTICS OF THE COST DATA

2.1. Introduction

The purpose of this section is to give some insight into the data used in the further analyses in this paper. As described above, the analysis focuses solely on the data in the CPPAP, which is calculated using the "Allocation Area Problem Ranking Worksheet."[1] This worksheet is shown in Fig. 1. Column B, the cost factors, are specific to the Prototype District, but they are also representative of other loop plant districts. Abbreviations used in Fig. 1 and throughout this section are as follows: LST—line and station transfer; WOL—wired out of limits; BCT—break connect-through; CDP—clear defective pair; BPC—break permanent connection; CIR—control point interconnection; RE—referred to engineer; RTC—reterminated connection; AC-SOD—assignment change because the originally assigned pair from a service order was found to be defective; AC-NS—non-service-order assignment change; AC-OTH—other assignment change; FCT-7AB—7A or 7B cable trouble associated respectively with splicing and terminating troubles; FCT-OTH—other cable trouble; DEF PRS—defective pairs. For definitions and discussion of these and other loop plant terms, see Ref. 4.

Two of the items on the worksheet were not measured directly in the data base. They are the BCT and RTC. However, based on engineering studies in the Prototype District[6] it was determined that these could be adequately approximated for the Prototype District during the study period by a fraction of the total facilities assigned, which is measured in the data base. These studies determined that BCTs were 13 percent of the facilities assigned and that RTC were 35 percent of facilities assigned. Finally, the management of the loop plant used in the Prototype District was such that there were no CDP, BPC, or CIR. Therefore, in all further analyses these cost factors are ignored. All other variables, except the number of defective pairs, are available (monthly) in the data base. Defective pairs were entered annually from the district's yearly pair status report. This report gives the pair status (e.g., assigned, defective, etc.) as of January 1 and is used monthly for the twelve month period centered at January 1 (i.e., July through June). Thus, the data to be studied in this section are the monthly values of the CPPAP and the 11 sub-components of CPPAP that were either measured or estimated during the study.

AA ______________ DATE OF RANKING ______________

LINE #	ITEM	A ENTRY	B COST FACTOR	C COST PENALTY
1	LST	______ #/YR	X 17.52	= ______
2	WOL	______ #/YR	X 36.81	= ______
3	BCT	______ #/YR	X 7.64	= ______
4	COP	______ #/YR	X 72.70	= ______
5	BPC	______ #/YR	X 24.64	= ______
6	CIR	______ #/YR	X 70.55	= ______
7	RE	______ #/YR	X 35.15	= ______
8	RTC	______ #/YR	X 9.48	= ______
9	AC - S. O. Def	______ #/YR	X 29.35	= ______
10	AC - Non S. O. Def	______ #/YR	X 68.14	= ______
11	AC - Other	______ #/YR	X 32.63	= ______
12	FCT - 7A, B	______ #/YR	X 83.32	= ______
13	FCT - Other	______ #/YR	X 109.00	= ______
14	Def Prs	______ #DEF Pr X____ DIST CO KFT	X 0.91	= ______
15	TOTAL COST PENALTY (SUM 1 TO 14)			______
16	COST PENALTY PER ASSIGNED PAIR ______ ÷ ______ # ASSIGNED PAIRS LINE 15			= $\boxed{}$

Fig. 1—Allocation area problem ranking worksheet.

2.2. Components of CPPAP

The CPPAP has 11 non-zero cost components. However, two of those variables are perfectly correlated since they are both proportions of the facilities assigned (i.e., BCT and RTC). Therefore, since both the cost factor (see Fig. 1) and the proportion of facilities assigned associated with the RTC is higher than that for BCT, it is the RTCs which will be used in the further analyses in this subsection. In later sections of the paper all components are used in the calculation of CPPAP.

A numerical summary of the level (mean) and variability (standard deviation) of the ten cost components for each of the three stages of area conversion is given by Table I. So that a few extreme data values do not overwhelm the rest of the data, the 25 percent trimmed mean and standard deviation were used. Thus these values are based on only the

Table I — CPPAP component costs for 10 converted areas

Variable	Trimmed mean			Trimmed std dev		
	Before	During	After	Before	During	After
LST	0.65	0.12	0.0	0.77	0.17	0.0
WOL	0.04	0.0	0.0	0.18	0.0	0.0
RE	0.77	0.04	0.0	1.20	0.19	0.0
RTC	0.53	0.40	0.21	0.22	0.14	0.09
AC-SOD	0.26	0.10	0.01	0.38	0.18	0.12
AC-NS	1.44	3.78	0.46	1.16	3.35	0.52
AC-OTH	1.12	1.47	0.44	0.56	1.41	0.41
FCT-7AB	2.30	4.97	0.21	1.36	4.50	0.61
FCT-OTH	0.19	0.13	0.0	0.82	0.39	0.0
DEF PRS	0.80	0.84	0.85	0.63	0.83	0.89

middle 50 percent of the data. First the trimmed mean across months for each area in each stage of conversion was computed; the tabled values are the trimmed mean and trimmed standard deviation of those values across the 10 converted areas. Focusing on tbe mean (level) values first, it is clear that the dollar costs shown in the table vary widely from component to component as well as for the stages of conversion. Perhaps the most remarkable change is in the non-service-order assignment change tickets (AC-NS) which go from \$1.44 before to \$3.78 during to \$0.46 after. However, considering the physical situation, this type of behavior is to be expected. During the conversion, many of the cable pairs are being handled by the nature of the design of an allocation area. This can cause many of the pairs to become defective and can cause an interruption in the customer's service. The service is restored either by changing the customer to a new pair (recorded as an AC-NS) or actually fixing tbe defective pair (recorded as an FCT-7AB). Note further that the occurrences of splicing and terminating cable troubles (FCT-7AB) also peak during conversion and fall to greatly reduced levels in the after period. However other cable troubles (FCT-OTH) contribute little to CPPAP. The category of assignment changes due to the originally assigned pair from a service order being defective (AC-SOD) drops to very nearly zero after conversion. Other assignment changes (AC-OTH) is a major contributor to CPPAP during all three periods of conversion. The LST, WOL, and RE after conversion all have zero trimmed mean and standard deviation. The category of defective pairs (DEF PRS) is interesting because its level stays the same from during to after, and its variability actually increases during this transition. However, since the defective pair data is only updated annually, these results should be considered preliminary. More detailed special studies of defective pair rates have been performed and are included in Ref. 2.

While the table is a helpful summary of overall behavior, it is not useful in trying to characterize the similarity and differences among the areas with regard to the components of CPPAP. Graphical displays of multivariate data are often useful for gaining insight into the basic structure

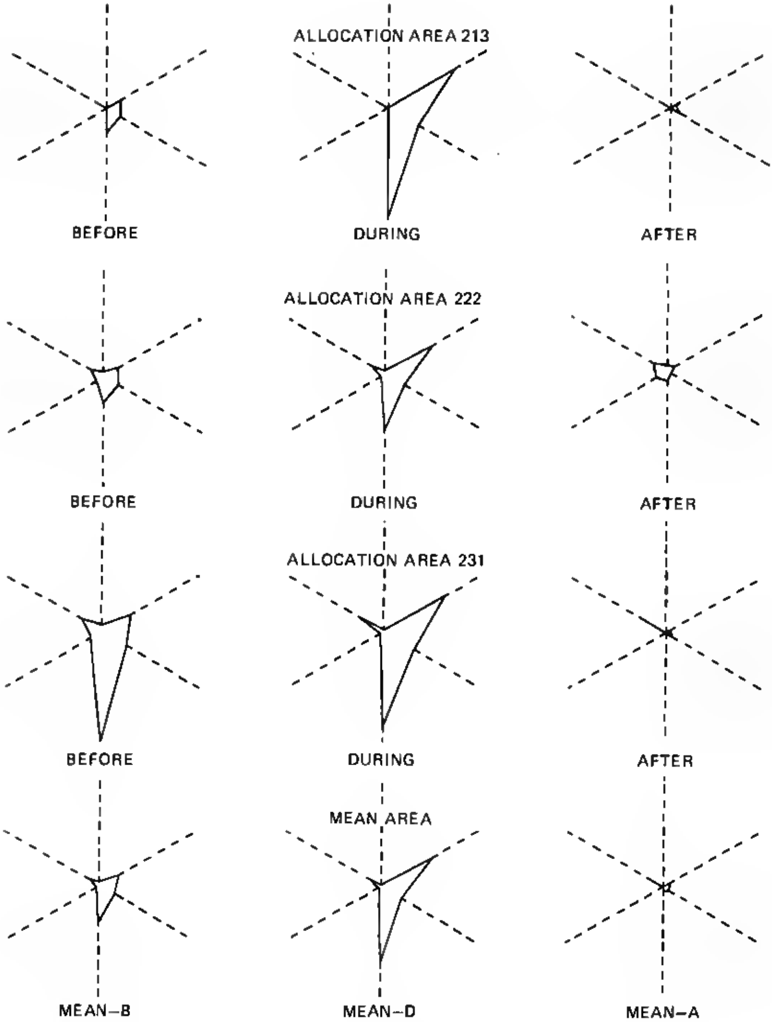
of data. However, they tend to become more complicated and less useful as the number of variables increases. Based on Table I, it seems fairly clear that most of the interesting (large and variable) dollar components of CPPAP are found in the assignment changes, the cable troubles, and the defective pairs. The costs associated with LSTs, WOLs, REs and RTCs tend to be both small and fairly stable. Therefore, in the graphical displays the focus will be on the six largest and most diverse cost components.

Figure 2 gives one example of a polygon plot[7] for three of the converted areas and the mean converted area (i.e., the 25 percent trimmed mean of the converted areas). The polygon is formed by connecting the value of each variable plotted on its respective axis (see Fig. 2 key). By examining the polygons associated with different areas and stages of conversion it is possible to visually compare and contrast characteristics of the areas. Note the similarity of the areas for before, during, and to some extent after. The values in these plots are as in Table I, and show dollar amounts. The scaling is designed to show most of the variability in these data without being distorted by a few very large values. Although areas of a polygon do not directly correspond to the total cost associated with an allocation area, areas do give some idea of that sum. For example, it is clear that after conversion the cost penalty is very small compared with during and before. The anomalous large value of the non-service order assignment changes (mentioned earlier) is evident in the during period. The peak on the first axis from the vertical position is this large value.

2.3. Analysis of CPPAP

To achieve an initial feel for the nature of the CPPAP data, a plot of these values against time for the individual allocation areas is useful. Figure 3 shows a sequence of four allocation areas for their entire 33 month data history. Note that the vertical scales on the four plots, which show dollar cost penalties, are different. While such differences make across area comparisons difficult, the range of the data (particularly including converted and non-converted areas) is so large that using a single scale would obscure much of the available detail. Because there is a good deal of variability in the CPPAP measure, a non-linear (resistant) smoother is applied to the data and plotted (as the solid line) along with the raw values. The resistant smoother used is (3RSR), twice.[8] Since this smoother is based on moving medians, rather abrupt changes may occur in the smoothed output. This smoother was selected for just this reason so that rapid changes in the level of the data (e.g., after conversion) would not be obscured.

Of these four allocation areas (212 through 215), two were eventually converted (213 and 214), while the other two were not. For those areas which have been converted, lines are drawn to indicate the end of the



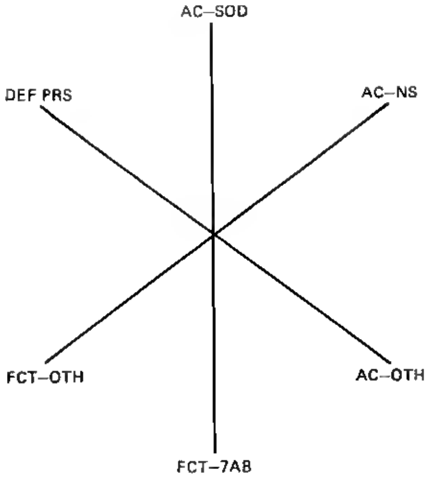
(a)

Fig. 2—(a) Components of CPPAP (radius length 7.12). (b) Key.

before conversion period, and the beginning of the after conversion period. Note that these vertical lines are drawn between actual monthly observations. The data accuracy only allows full month designations of before, during, or after. For example, in area 214 months 1–5 are before, months 6–13 are during, and months 14–33 are after conversion.

Analysis of this figure (and others) showing all the area-time histories gives a considerable amount of insight into the nature of the data.

(*i*) The CPPAP for the areas where there is no conversion tends to be more stable than for areas that undergo conversion.



(b)

Fig. 2. (*continued*)

(*ii*) Fairly large excursions from a smooth value are evident for all areas. (Note that the resistant smoother is not affected by these unusual excursions.)

(*iii*) The level and variability of the before, during, and after may be quite different.

(*iv*) The after conversion behavior of these areas is quite different. For example, in area 213 the CPPAP drops quickly to a value near zero. In area 214 there is a slow but steady decline to a near zero value for CPPAP.

(*v*) No evident seasonal pattern is visible in this limited amount of data.

Table II shows a basic summary of the behavior of each of the 10 converted areas for before, during, and after conversion months. The 25 percent trimmed mean and standard deviation are used, as in Table I, so that the tabled values reflect the bulk of the data. Table II shows that both the level and variability change during the "life" of an area. The during period tends to have the highest levels. The after is the lowest (as would be both expected and presumed because the effect of conversion is to reduce the occurrence of the costly plant troubles) both in level and variability. The variability of the before conversion data is quite high and not uniform across areas.

In summary, based on these and similar displays, CPPAP values appear to vary quite widely both across allocation areas and stages of conversion. For those areas which were converted, the level and variability of the individual components of CPPAP tend to be concentrated in the assignment changes, cable troubles, and defective pairs.

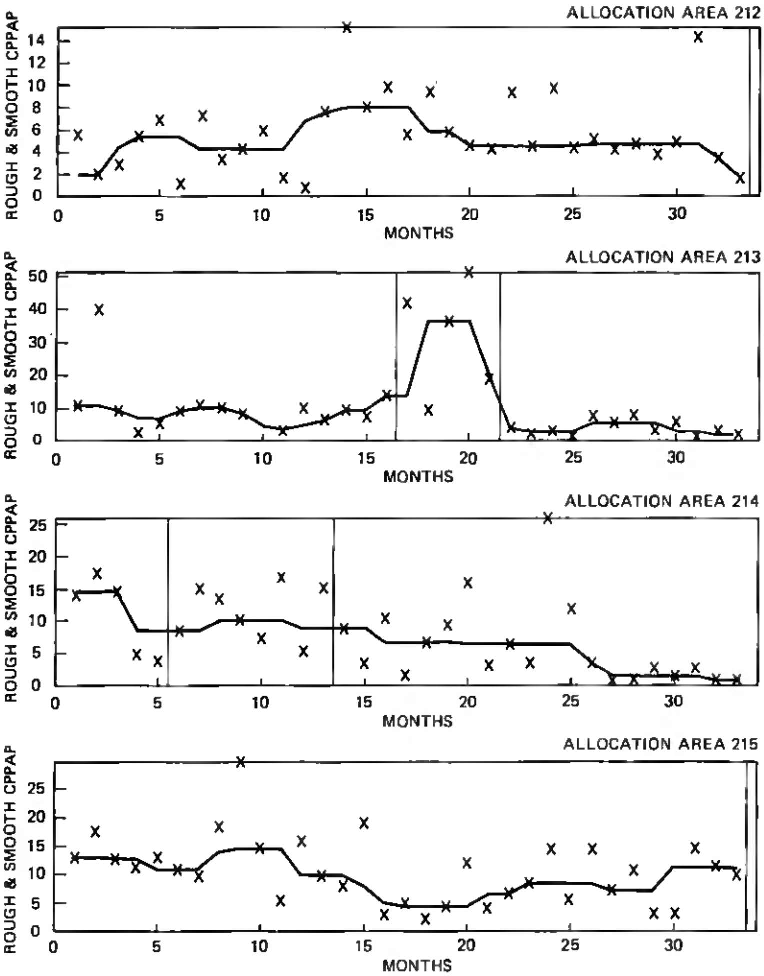


Fig. 3—Rough and smooth CPPAP for various allocation areas.

III. EXPLORATORY AND GRAPHICAL ANALYSES OF BEFORE CONVERSION DATA

3.1. Motivation

One use of the Prototype District Data Base is to develop methods of analysis for determining which allocation areas in other districts should be converted to SAC, using FAP techniques. Related questions concern how many months of data should be collected before making such decisions, and how large the areas should be in the first place so reliable decisions can eventually be made. This section explores certain properties of these data, motivated by these goals. Fluctuations in the cal-

Table II — CPPAP for 10 converted areas

Area	Trimmed mean Before	Trimmed mean During	Trimmed mean After	Trimmed std dev Before	Trimmed std dev During	Trimmed std dev After
209	4.39	6.68	1.90	2.84	5.20	1.10
210	5.97	22.32	3.51	6.52	10.98	2.22
213	8.84	32.56	3.35	3.40	31.65	3.25
214	11.43	11.81	4.18	14.75	7.73	5.98
221	9.58	10.29	1.91	6.86	9.58	0.00
222	9.41	11.64	4.01	3.02	5.03	0.00
227	10.37	10.07	6.67	6.85	4.30	4.95
228	11.73	14.97	4.22	4.91	7.61	1.19
229	14.15	17.00	3.02	10.86	22.88	0.93
231	21.10	17.03	3.32	7.52	7.28	2.04

culated cost penalty across months and across areas can be large, as was seen in Section III. Thus, statistical methods are needed to help answer these questions. Since only before conversion data could be used to help in making decisions regarding conversion, only the before data from the data base are considered here. The analysis uses the cost measure CPPAP for reasons described in Section I.

The goal here is to examine the structure of the before conversion data so as to be led to reasonable methods of analysis (i.e., reasonable assumptions and models) to answer these questions. We concentrate on searching for and examining certain relationships by studying appropriate scatter plots. While certain numerical statistics are also useful for such purposes, an advantage of plots is that they are more exploratory in nature. Section IV then presents and uses a specific model, supported by the data, as a way of answering the questions in the previous paragraph.

3.2. Analyses

For the following plots, consider the cost penalty x_{ij} for area i and month j. The mean, $\overline{x}_i$, and standard deviation, s_i, of these values across months for each area were calculated. Only before conversion data were used, so the number of months differs from one area to another; however, recall that 13 of the 23 areas were never converted, so for these areas all 33 months are available. Figure 4 plots the standard deviation s_i vs. the average cost penalty $\overline{x}_i$ for all 23 areas. A positive relationship between these two quantities is very clearly apparent. Such a relationship strongly violates assumptions that would be desirable and convenient to use.

Another look at this relationship can be obtained by considering the sizes of the 23 areas. Since the cost penalty x_{ij} is itself an average calculated over the number of pairs in the area (cost penalty per assigned pair), one might expect the standard deviation of these values, s_i, to be smaller the larger the size of the area. Figure 5 plots s_i vs. the number of assigned pairs in the ith area, p_i. From theoretical grounds one might expect the relationship between s and p to be of the form $s = \sigma/\sqrt{p}$, for

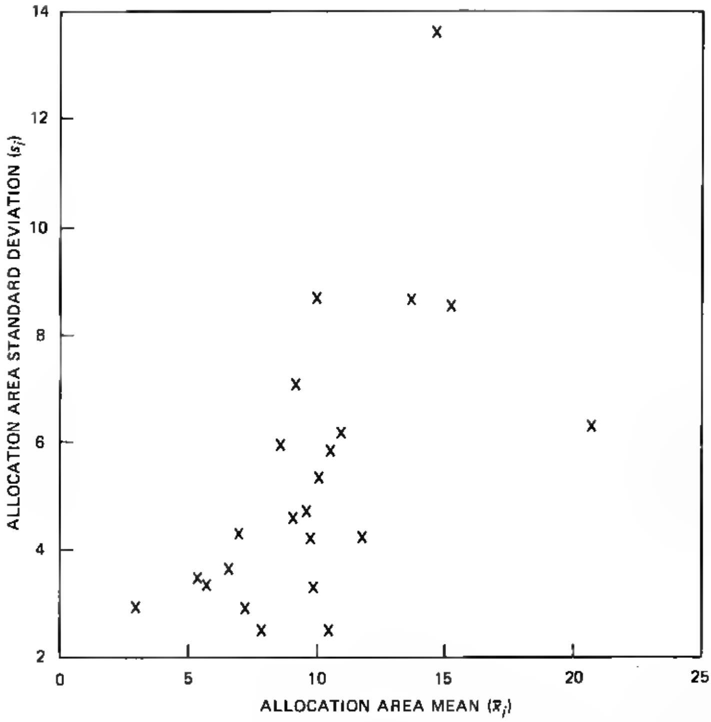


Fig. 4—CPPAP values before conversion.

some σ. The points in Fig. 5 look like they might generally follow a relationship like this, plus some scatter. Thus, we fit a curve $\hat{s} = \hat{\sigma}/\sqrt{p}$ to these points using least squares* and then formed the residuals $(s_i - \hat{s}_i)$. Each residual is plotted against the corresponding $\overline{x}_i$ in Fig. 6. Again a strong increasing relationship is apparent; the larger the average cost penalty $\overline{x}_i$ for an area, the more likely it is that $(s_i - \hat{s}_i)$ is positive and large. Even after removing the effect of area size from the standard deviation s_i, higher area averages $\overline{x}_i$ are associated with higher area standard deviations s_i.

One approach to answering the questions put forth in Section 3.1 would be to fit an appropriate linear statistical model to these data, and then make inferences from that model. However, one of the assumptions underlying the usual fitting of such a model is that of homogeneity of variance; i.e., the variance of the observations should be constant across different levels of other variables. Because of the relationships seen above, it is worthwhile also to consider transformations of CPPAP when exploring the before conversion data. Some transformed variable quite possibly could be generally appropriate for later, more formal analysis than would the raw CPPAP values.

* Weighted least squares were used, for reasons described below.

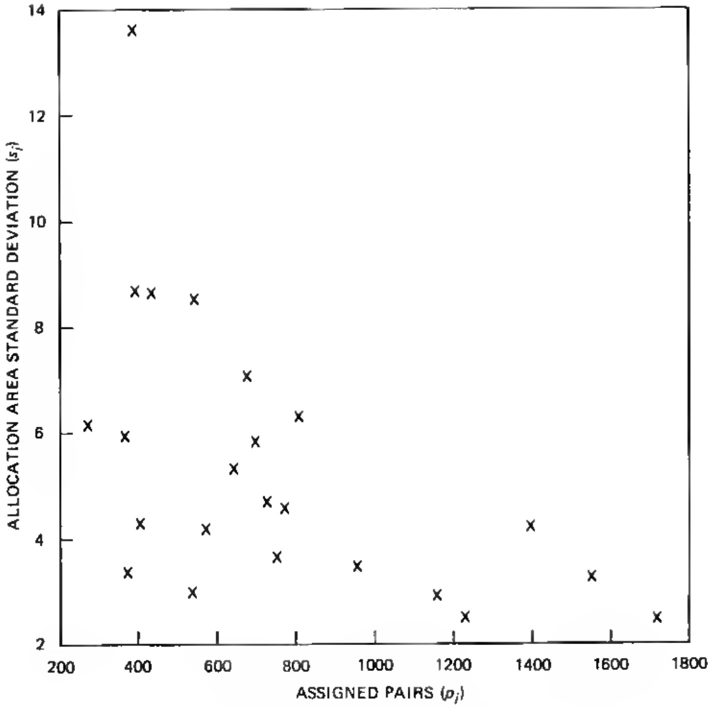


Fig. 5—CPPAP values before conversion.

Several transformations of the cost penalties within the family $y = (x + a)^b$, with a and b specified parameters, were calculated and studied. Considering the results as a whole, the most satisfactory and interesting properties appeared using the transformation $y = \ln(x + 1)$, which corresponds to $b = 0$, with x the CPPAP as before. Thus the following plots in this section were all constructed using this transformation.

Figure 7 plots the standard deviation $(sy)_i$ vs. $\bar{y}_i$, with the plotting character showing the size of the area; "1" for areas with assigned pairs $p_i \leqslant 500$; "2" for $500 < p_i \leqslant 700$; "3" for $700 < p_i \leqslant 950$; "4" for $p_i > 950$. There appears to be no systematic relation between (sy) and $\bar{y}$, although the two extreme (high and low) values on $\bar{y}$ possibly suggest a decreasing trend; certainly there is nothing like the behavior in Fig. 4. Moreover, the higher number plotting characters tend to be at the bottom of the plot with the lower numbers at the top, implying that larger areas have smaller variability, apart from their average value. The area average $\bar{y}_i$ is plotted against size p_i in Fig. 8; these quantities appear unrelated, so knowing *a priori* the size of an area does not enable one to say much about its expected average cost penalty.

Figure 9 shows the standard deviation $(sy)_i$ plotted against size p_i. There is a downward trend, and one expects larger areas to have smaller

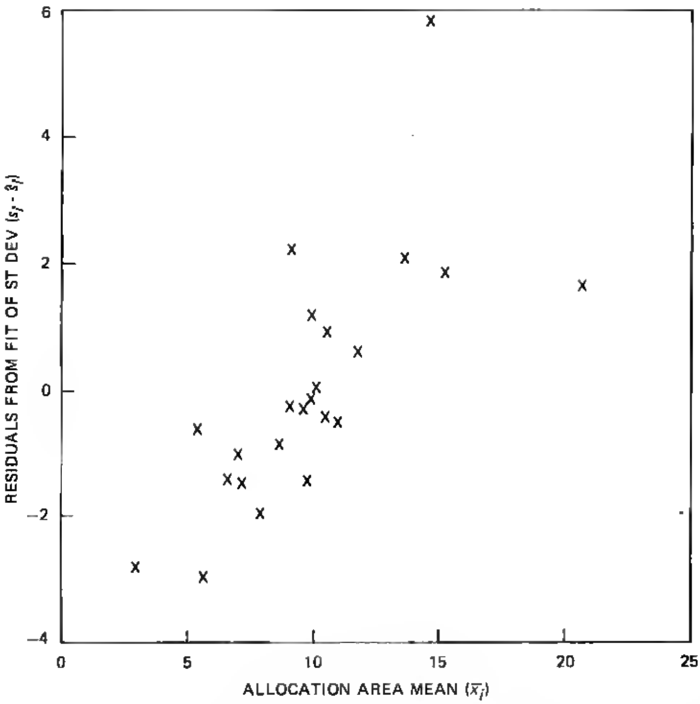


Fig. 6—CPPAP values before conversion.

variability. To see to what extent this trend is accounted for by a $sy = \sigma/\sqrt{p}$ relationship, $\hat{\sigma}$ was obtained by a weighted least squares regression of $(sy)_i$ on $1/\sqrt{p_i}$; the fitted curve is the solid line in Fig. 9. A weighted regression was used because the variances of the individual points $(sy)_i$ about their expectations $\tau_i = \sigma/\sqrt{p_i}$ depend on the values of τ_i and m_i, the number of months of before data for that area; assuming normality of the y's, the variance is $0.5{\cdot}\tau_i^2/(m_i - 1)$. (This is derived from the χ^2 distribution associated with $(sy)^2$.) Thus, weights proportional to the reciprocal square roots of these variances were used, and the following three plots are the raw residuals multiplied by these weights.

The residuals $sy_i - \widehat{sy}_i$ are plotted against $\bar{y}_i$ in Fig. 10. No strong relationship is apparent. Perhaps the points with extremely high and low $\bar{y}$ suggest a downward trend, but if these single points are ignored no structure at all remains. Figure 11 plots each residual against m_i, the number of months of before data for that area. One would like to see a horizontal band, which would signify no relationship; indeed, the plot does not suggest any strong relationship. A normal quantile-quantile probability plot[9] of the residuals is displayed in Fig. 12. This shows reasonably good normality of the residuals, although the largest value is somewhat larger than would be expected and there is some bunching of the residuals, for which we have no explanation.

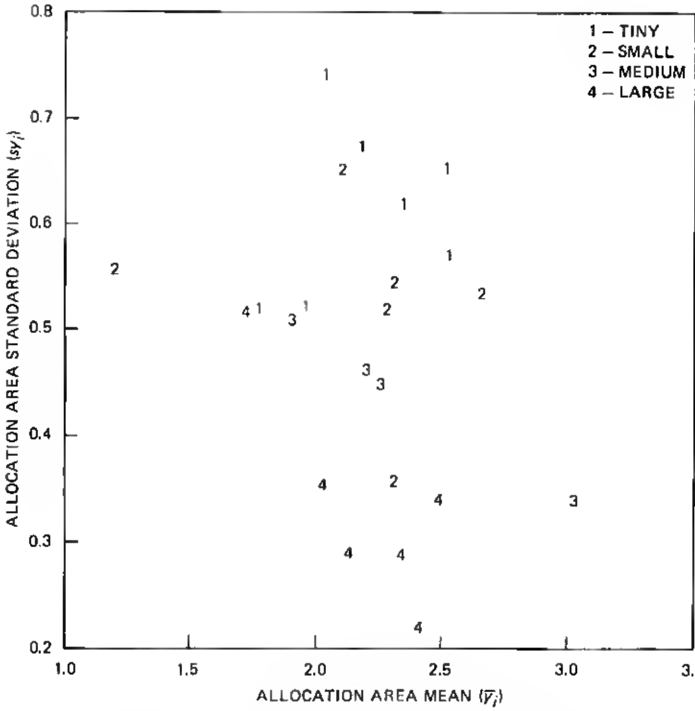


Fig. 7—Values of ln(CPPAP + 1) before conversion.

Thus, for the logarithmic transform of the original cost penalty nicer behavior results than with the raw variable. An area's standard deviation is unrelated to its level, but it is related to its size in a reasonable way; moreover, the residuals from this relationship have reasonable properties. A number of additional properties of these data were explored, but to conserve space only a few will be discussed in any detail.

For each month, the mean and standard deviation of the CPPAP values for all allocation areas for that month were calculated. Figure 13 plots the monthly standard deviations vs. the monthly average, again using $y = \ln(\text{CPPAP} + 1)$. There are 33 points in the plot, one for each month; of course the points from later months are based on successively fewer values as areas are converted. No relationship is apparent; this is consistent with the lack of relationship between standard deviation and mean as calculated for each area in Fig. 7. The monthly average vs. the month number and the smooth of these data [using 4(3RSR)2, twice, a non-linear smoother[8]], are shown in Fig. 14. This suggests somewhat of a cyclic behavior in the average cost penalty. Local peaks appear around months 1–2, 12–14, and 26–28. One might hypothesize the existence of a cyclic 12-month structure to these data due to seasonal local factors such as weather, churn, and inward and outward movement. However,

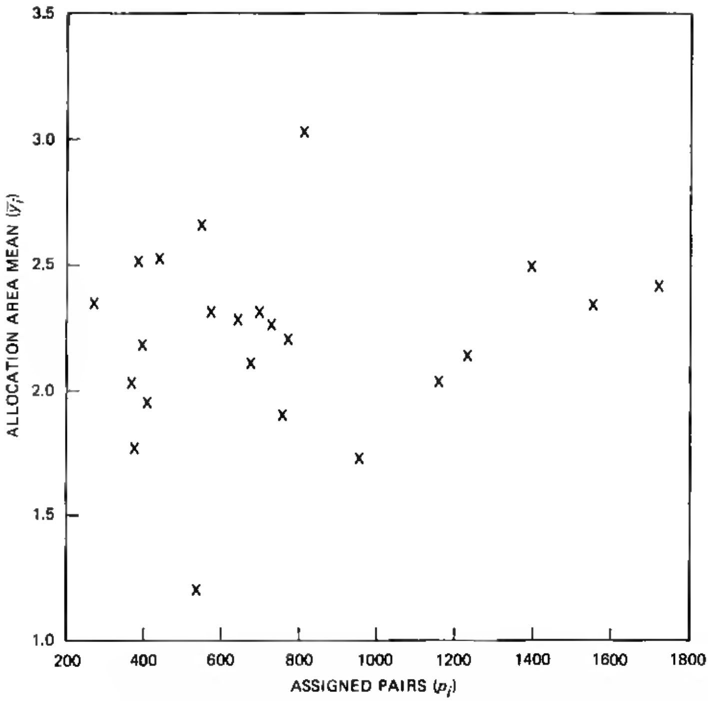


Fig. 8—Values of ln(CPPAP + 1) before conversion.

Fig. 14 does not show such clear behavior that one could extrapolate some fitted cycle with any confidence. Moreover, recall that the purpose of these analyses is to develop methods that could be used with (probably less extensive) data from other districts for decision making. We would not want to extrapolate a specific seasonal pattern from Fig. 14 to a new district without careful consideration of similarities and differences between the new district and the Prototype District. One might, though, wisb to use 12 or 24 months data when arriving at decisions so as to remove seasonal effects. The possible seasonal factor is discussed further in reference to somewhat different purposes in Section V.

Distributional characteristics and the correlation structure of the transformed observations can also be of interest. Figure 15 gives a normal quantile-quantile plot of $(y_{ij} - \bar{y}_{i\cdot})\cdot\sqrt{p_i}$ for all areas i and months j before conversion. This quantity is of interest because some differences between areas are expected, but can be removed by looking at the deviations $y_{ij} - \bar{y}_{i\cdot}$. No strong monthly effect was seen above, so that possibility is ignored here; and also it was found earlier that $\text{var}(y_{ij})$ is approximately σ^2/p_i, so the values $(y_{ij} - \bar{y}_{i\cdot})\cdot\sqrt{p_i}$ should have approximately equal variance. Figure 15 shows that these values are distributed reasonably closely to the normal distribution.

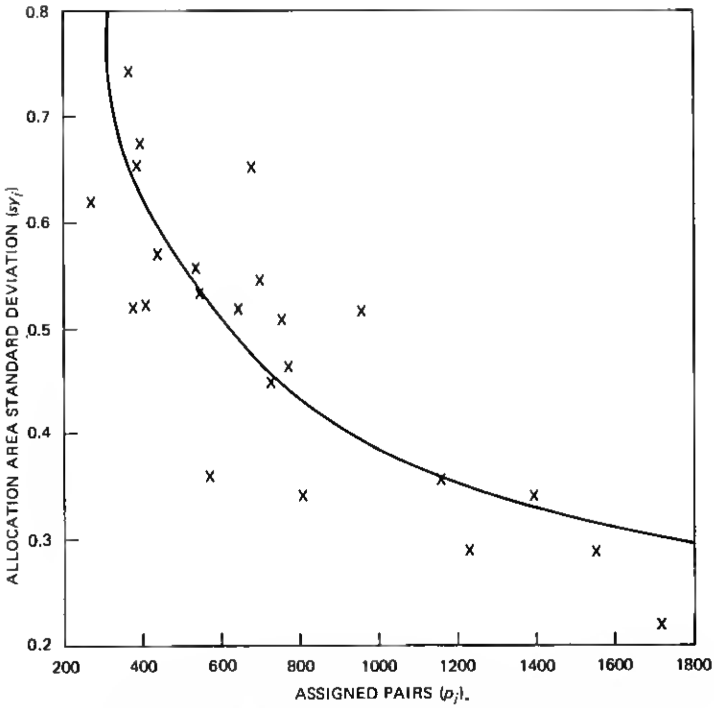


Fig. 9—Values of ln(CPPAP + 1) before conversion.

Turning to the possible relationships between areas, a different normal quantile-quantile plot, calculated from correlations in the following way, is given in Fig. 16. For each pair of areas k and l, the correlation between the above $(y_{ij} - \bar{y}_{i\cdot})\cdot\sqrt{p_i}$, $i = k$ and l, was calculated over the before conversion months common for both areas. This gives 253 (= 23·22/2) estimated correlations, and we would like to see to what extent these differ from a random sample of correlations where the true correlation coefficient is 0. Fisher's z transformation,

$$z = \tfrac{1}{2}\ln\left(\frac{1+r}{1-r}\right)$$

was used to achieve approximate normality. If the population correlation is 0, then mean $(z) \approx 0$,

$$\mathrm{var}(z) \approx \frac{(n+1)}{(n-1)^2}$$

where n is the sample size and z is approximately normally distributed. For these data each z was divided by the standard deviation corresponding to the number of months n from which it was calculated, and Fig. 16 is a normal quantile-quantile plot of the standardized z's. A

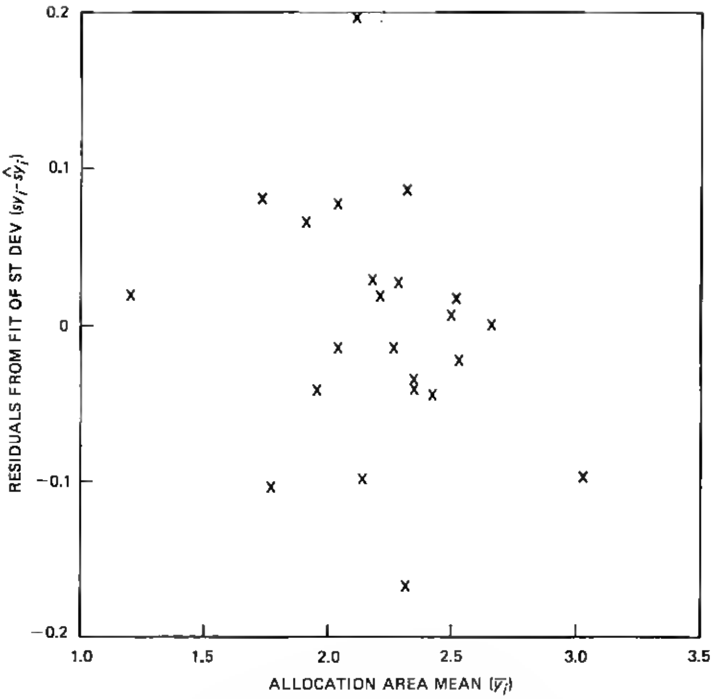


Fig. 10—Values of ln(CPPAP + 1) before conversion.

"perfect" result would have all points on the $y = x$ line, which is drawn on the plot. However, even if the true correlation were 0 one would not necessarily expect our standardized z's to scatter exactly about this line since we do not have 253 correlation coefficients calculated independently of one another. Instead they are formed pairwise from 23 variables, implying some (complicated) structure among them. In Fig. 16 the points are uniformly above, but quite close to the $y = x$ line; the standardized z's are slightly but consistently larger than would be expected if all true correlations were 0. The median of the standardized z's corresponds to a population correlation of about 0.3. Thus there is evidence of a positive but not large correlation between the values in different areas at the same point in time. This result is not intuitively unexpected since geographic proximity is probably the cause. For example, a heavy rainstorm may increase cable troubles and hence larger values of CPPAP. A more exhaustive exploration of the correlation structure of these data could also consider correlations both between and within areas at different points in time, i.e., with leads and lags.

Another plot of some interest, Fig. 17, shows $\bar{y}_i$ vs. the distance of each area from the central office, d_i. Although one might or might not expect such a relationship, the data strongly suggest that areas further from the

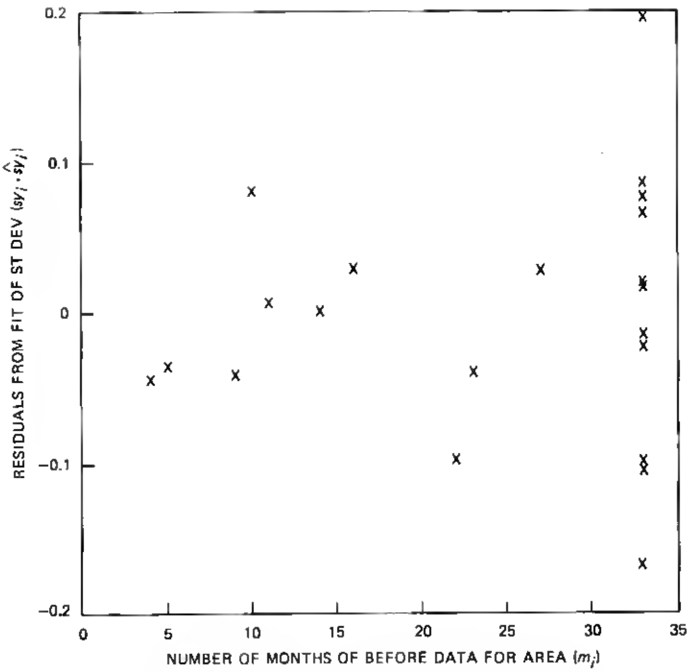


Fig. 11—Values of ln(CPPAP + 1) before conversion.

central office have higher cost penalties. It would be of interest to have explanations for this and to see if this relationship generalizes to other districts. Such investigations are in progress by the authors and others. However, as with the possible monthly cycle seen above, we would not necessarily want to extrapolate this in a straightforward way to other districts. It is also of interest to consider the plot of the weighted residual $(sy_i - \hat{sy}_i)$ vs. d_i, given in Fig. 18. Although the area average may be related to d_i, Fig. 18 shows that the part of the standard deviation not predicted from the size of the area does not seem related to d_i. This latter result fits in with the previous discovery that the standard deviations of the y's do not appear to be systematically related to anything except the size of the area.

The entire set of plots and analyses described in this section were repeated using robust estimates of location and scale instead of the sample mean and standard deviation. The purpose was to see if a small number of deviant observations might be either causing, or hiding, the relationships considered above. However, there was no appreciable difference in the results. The results using the mean and standard deviation, rather than the more robust statistics, were presented above because of the widespread familiarity and use of these statistics.

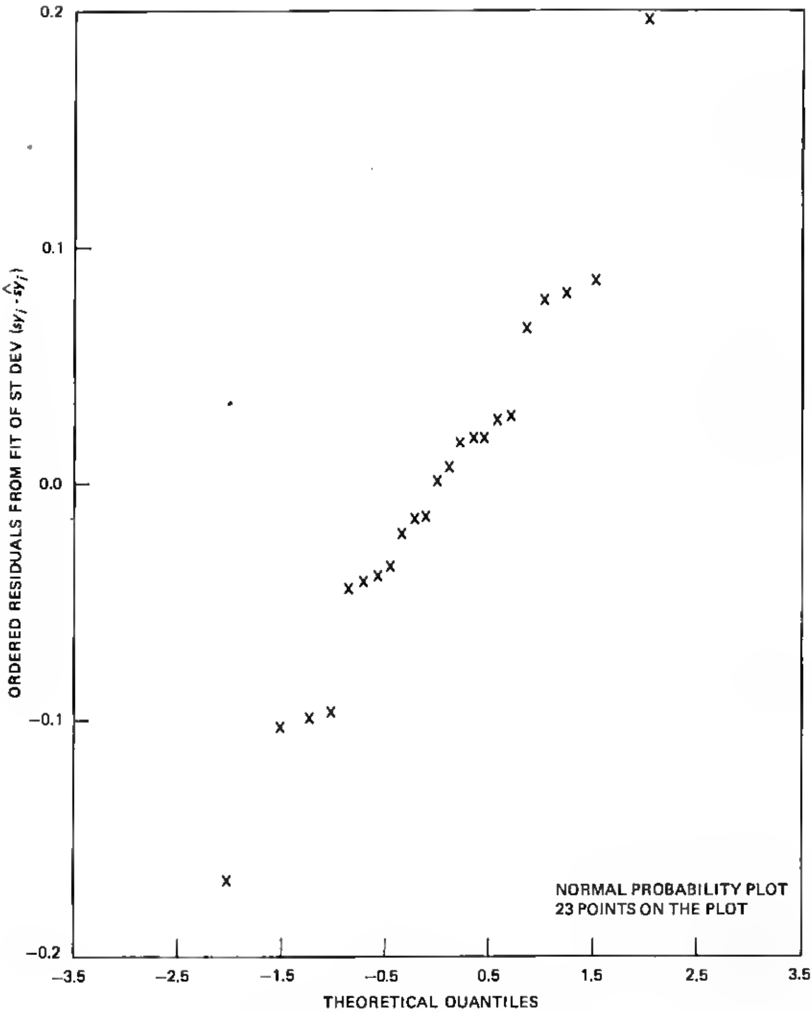


Fig. 12—Values of ln(CPPAP + 1) before conversion.

The analyses were also repeated using other cost measures. As in the case of CPPAP, for each of these measures some transformation of the original values was discovered which appeared more useful for interpretation and later analysis than was the raw cost measure.

IV. DATA COLLECTION GUIDELINES

4.1. General results

This section makes use of the results from the previous section to construct guidelines for the collection period and size of future allocation areas. These guidelines are in the form of confidence intervals for the

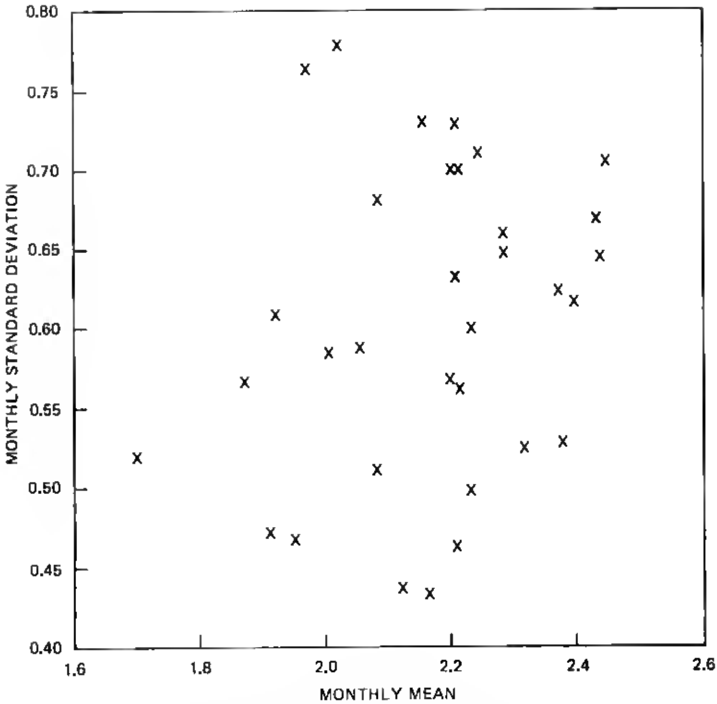


Fig. 13—Values of ln(CPPAP + 1) before conversion.

"true savings" given estimated savings, size of area, and the number of months of data collection. In addition, methods are presented for extending these results to local areas with characteristics different from those of the Prototype District.

Based upon the data analysis of Section III, it is reasonable to use the following model and analysis. Let $y_{ij} = \ln(\text{CPPAP} + 1)$ be the transformed cost measure for area i and month j. Express this as

$$y_{ij} = \mu_i + e_{ij} \tag{1}$$

where μ_i is the "true transformed CPPAP" for this area and e_{ij} is the "error" term corresponding to this month. We wish to make inferences about the area values μ_i and differences $\mu_i - \mu_j$.

Consider assumptions one can reasonably make concerning the e_{ij}. From theoretical grounds it is reasonable to assume that

$$\text{var}(e_{ij}) = \frac{\sigma^2}{p_i} \tag{2}$$

where p_i is the size, in assigned pairs, of the area. The quantity σ^2 can be interpreted as the inherent variability from one assigned pair in one month, and the error term e_{ij} results from averaging over p_i assigned

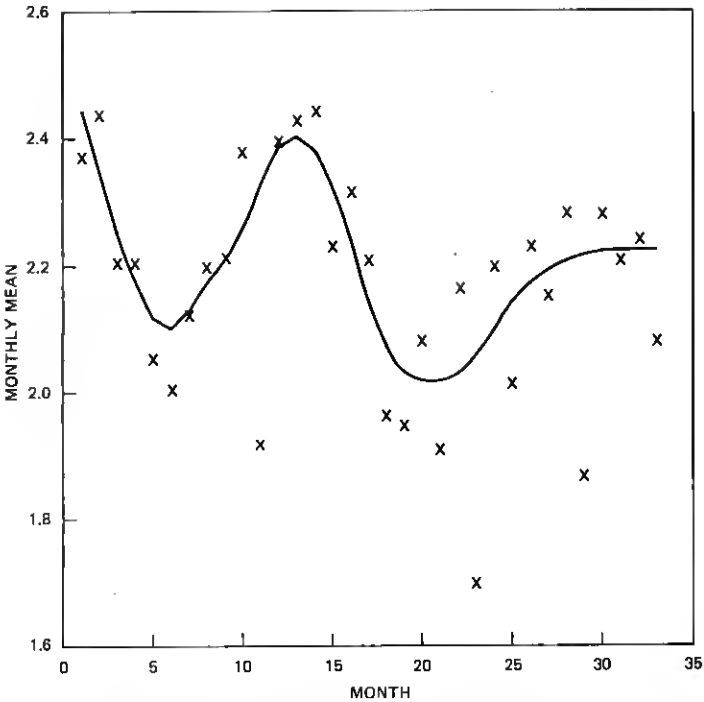


Fig. 14—Values of ln(CPPAP + 1) before conversion.

pairs. This assumption was supported by the analysis of Section III. Moreover, that analysis showed that the standard deviation (of the transformed CPPAP) does not seem to be related to any other available variable.

Considering further assumptions concerning the distribution of the e_{ij}, it would be convenient, natural, and relatively simple if we could assume that the e_{ij} are independently normally (Gaussian) distributed with 0 mean (and variance from eq. (2)). In support of these assumptions, it was shown in Section III that $\sqrt{p_i}\cdot$ the estimated e_{ij} (i.e., $(y_{ij} - \overline{y}_{i\cdot})\cdot\sqrt{p_i}$) were normally distributed after transformation. As for the independence assumption, these values were found in Section III to have a positive, although not extremely large, correlation between areas. However, the independence assumption between areas is important mainly for the confidence interval comparison of two different areas, as in eq. (5) below, and a positive correlation implies that that interval would tend to be conservative, i.e. longer than necessary.

Thus, for purposes of the analysis we assume that the e_{ij} are independent normal $(0,\sigma^2/p_i)$. Thus the estimate $\hat{\mu}_i$ in eq. (1) is $\overline{y}_{i\cdot}$; i.e., the "true transformed cost penalty" is simply estimated by the average of

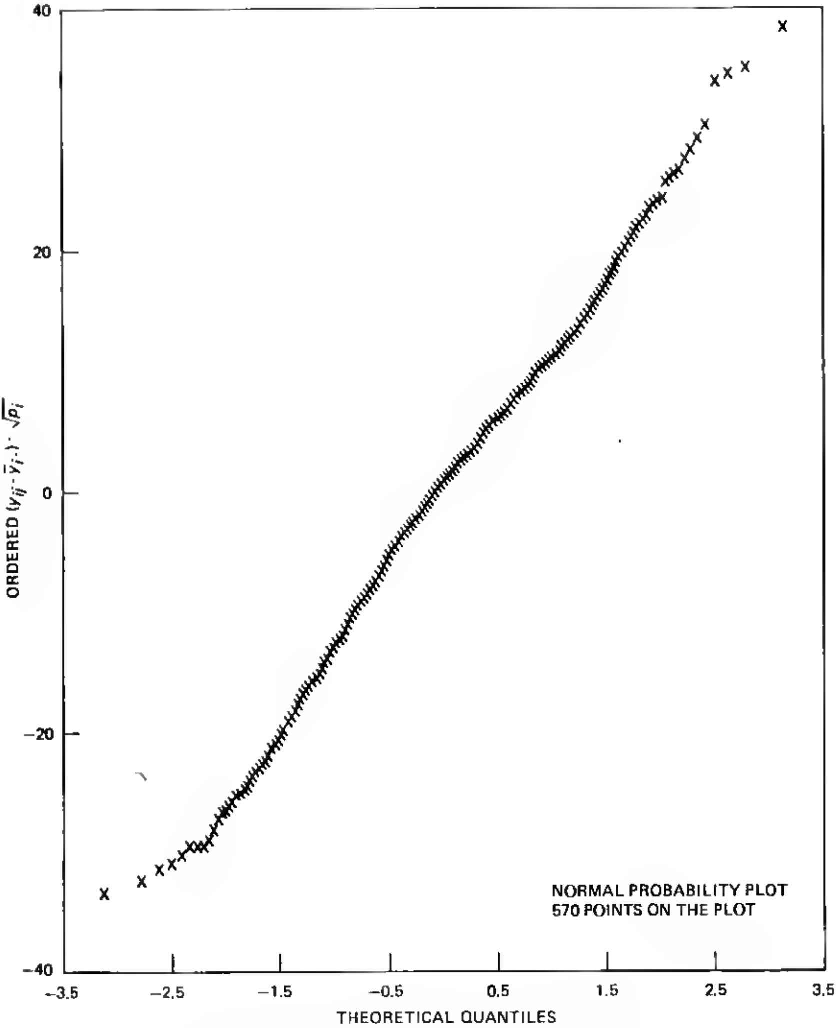


Fig. 15—Values of ln(CPPAP + 1) before conversion.

all observations for that area. Furthermore,

$$\operatorname{var}(\bar{y}_{i\cdot}) = \frac{\sigma^2}{p_i \cdot m_i} \tag{3}$$

where m_i is the number of months of before conversion values available for area i. Confidence intervals for μ_i (or $\mu_i - \mu_k$) can be calculated using eq. (3) and standard normal theory. A $100(1 - \alpha)$ percent confidence interval for μ_i is

$$\bar{y}_{i\cdot} \pm z \cdot \frac{\hat{\sigma}}{\sqrt{p_i \cdot m_i}} \tag{4}$$

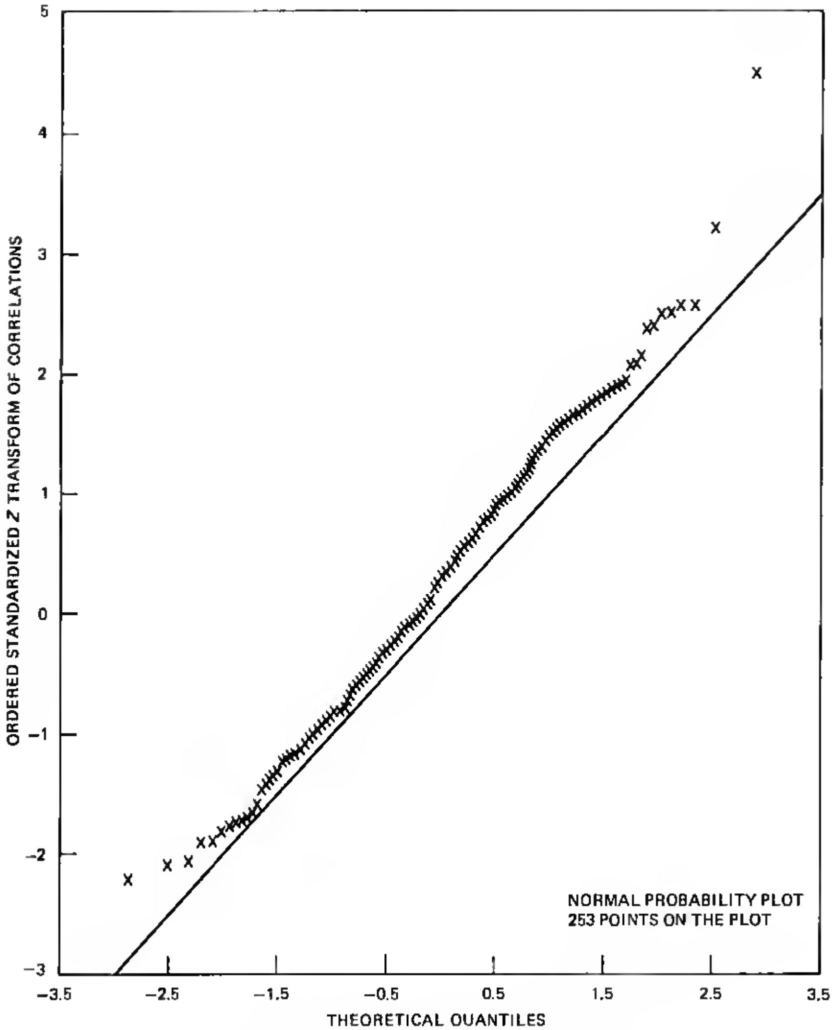


Fig. 16—Values of ln(CPPAP + 1) before conversion.

where z is the upper $1 - \alpha/2$ quantile of the standard normal distribution and $\hat{\sigma}$ is an estimate of σ described below. (Alternatively a t distribution could be used, but the degrees of freedom used in estimating $\hat{\sigma}$ should be large enough so that the difference in quantiles would be small.) Similarly, a confidence interval for the difference in "true" CPPAPs for two areas, $\mu_i - \mu_k$, is

$$(\overline{y}_{i\cdot} - \overline{y}_{k\cdot}) \pm z \cdot \hat{\sigma} \cdot \left(\frac{1}{p_i \cdot m_i} + \frac{1}{p_k \cdot m_k}\right)^{1/2} \tag{5}$$

The estimate of σ^2, $\hat{\sigma}^2$, is obtained from the regression

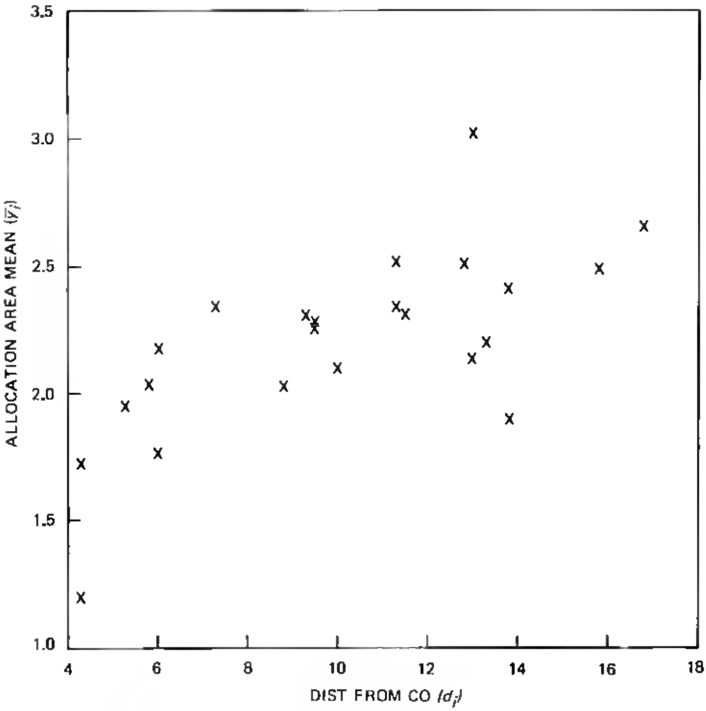


Fig. 17—Values of ln(CPPAP + 1) before conversion.

$$(sy)_i = \frac{\hat{\sigma}}{\sqrt{p_i}} + \epsilon_i \tag{6}$$

where $(sy)_i$ is the observed standard deviation of the m_i values in area i, and ϵ_i is an error term reflecting the departure of the observed $(sy)_i$ from this model. Eq. (6) is obtained from eq. (2) and its use is supported by Fig. 9 and other analysis in Section III. The variance of ϵ_i, given in Section III, depends on i, so an iterated weighted regression is performed to obtain $\hat{\sigma}$. Our value is 12.40. Thus the variance is effectively estimated by pooling results across all areas, while allowing for the fact that different sized areas have different variance.

Up to this point all the work in this section has been on variables measured on the transformed scale, i.e., ln(CPPAP + 1). Recall this transformation was selected to reduce the dependence of the variability on the level of CPPAP and to improve normality. Therefore, confidence intervals are for parameters μ_i, μ_k which are also transformed. However, we are interested in having tables (for example) based on the original data (untransformed) and representing untransformed parameters. This is simply done by forming the confidence intervals on the transformed scale and then performing the inverse transformation $x = e^y - 1$.

Shown in Table III are the 95 percent confidence intervals for various

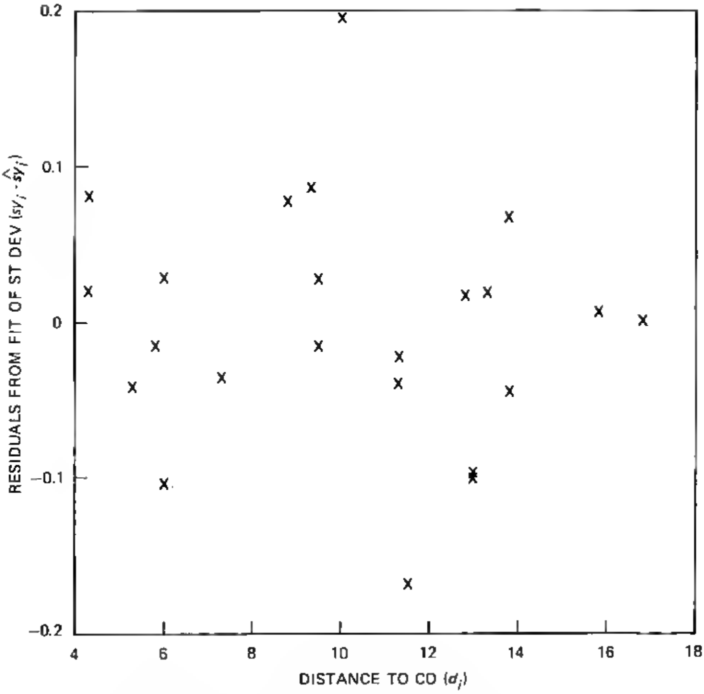


Fig. 18—Values of ln(CPPAP + 1) before conversion.

observed values of the CPPAP calculated using eq. (4) and $\hat{\sigma}$ estimated from the data. The time (in months) is the number of months used in forming the average value while the size is in pairs assigned. For example, suppose one has an area of 750 assigned pairs and has collected data for 12 months. If the computed average CPPAP is \$10, the confidence interval is from \$7.47 to \$13.29. If the computed CPPAP is \$30, the interval is \$22.87 to \$39.26. The interpretation is that 95 percent of the time, an observed CPPAP will be such that the associated interval covers the "true" CPPAP. Note that these intervals are not symmetric. On the transformed scale the assumptions yield a symmetric interval. However, when transforming back to the original scale, the nonlinearity of the exponentiation results in asymmetric intervals.

From the discussion of the variability of the average computed CPPAP it is clear that as the size of the area increases the variability decreases. Similarly, if the number of months used in computing the average CPPAP increases the variability of the estimate decreases. (In fact, based on eq. (4), and evident from Table III, the effects are symmetric.) To aid in assessing the magnitude of these effects Figs. 19 and 20 are provided. Figure 19 shows the upper and lower confidence limits for an observed CPPAP of \$20 formed by averaging over 12 months, for various values

Table III — 95% confidence intervals for:

Time (months)	Size (in assigned pairs) 250		500		750		1000		1250		1500	
	Cost penalty per assigned pair of 10 dollars											
3	3.45	26.20	4.80	19.87	5.52	17.55	5.99	16.30	6.34	15.49	6.60	14.92
6	4.80	19.87	5.99	16.30	6.60	14.92	6.99	14.15	7.26	13.65	7.47	13.29
9	5.52	17.55	6.60	14.92	7.13	13.88	7.47	13.29	7.71	12.90	7.89	12.62
12	5.99	16.30	6.99	14.15	7.47	13.29	7.77	12.79	7.98	12.47	8.14	12.23
15	6.34	15.49	7.26	13.65	7.71	12.90	7.98	12.47	8.18	12.18	8.32	11.98
18	6.60	14.92	7.47	13.29	7.89	12.62	8.14	12.23	8.32	11.98	8.46	11.79
	Cost penalty per assigned pair of 20 dollars											
3	7.49	50.94	10.07	38.84	11.45	34.42	12.35	32.02	13.01	30.48	13.51	29.39
6	10.07	38.84	12.35	32.02	13.51	29.39	14.25	27.92	14.77	26.96	15.17	26.27
9	11.45	34.42	13.51	29.39	14.53	27.40	15.17	26.27	15.62	25.53	15.96	25.00
12	12.35	32.02	14.25	27.92	15.17	26.27	15.75	25.33	16.15	24.71	16.46	24.26
15	13.01	30.48	14.77	26.96	15.62	25.53	16.15	24.71	16.52	24.17	16.80	23.78
18	13.51	29.39	15.17	26.27	15.96	25.00	16.46	24.26	16.80	23.78	17.06	23.42
	Cost penalty per assigned pair of 30 dollars											
3	11.53	75.67	15.34	57.81	17.38	51.29	18.71	47.75	19.68	45.48	20.42	43.86
6	15.34	57.81	18.71	47.75	20.42	43.86	21.51	41.70	22.28	40.28	22.87	39.26
9	17.38	51.29	20.42	43.86	21.92	40.92	22.87	39.26	23.54	38.16	24.04	37.38
12	18.71	47.75	21.51	41.70	22.87	39.26	23.72	37.88	24.32	36.96	24.77	36.29
15	19.68	45.48	22.28	40.28	23.54	38.16	24.32	36.96	24.87	36.15	25.28	35.57
18	20.42	43.86	22.87	39.26	24.04	37.38	24.77	36.29	25.28	35.57	25.66	35.05

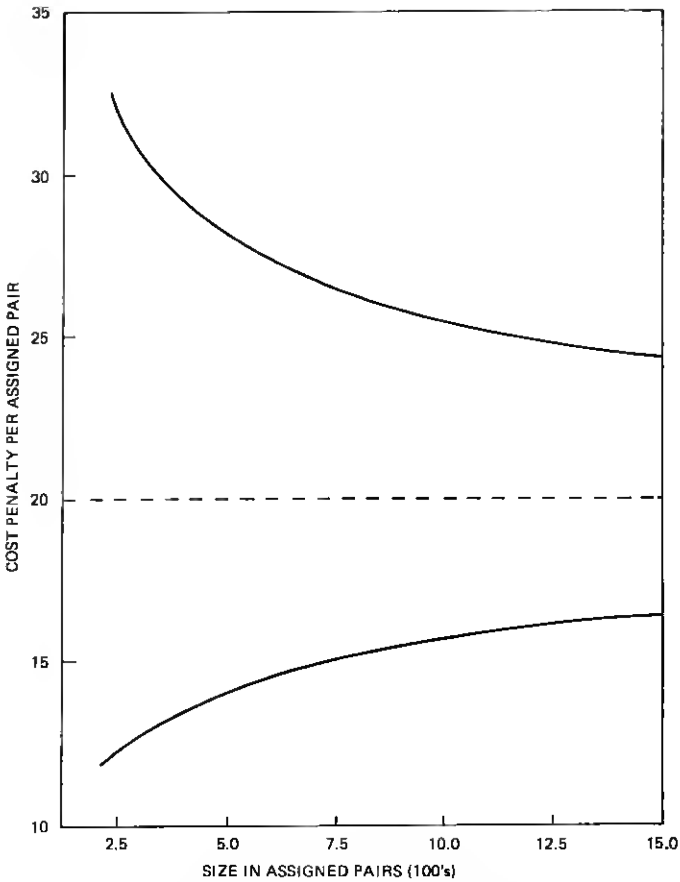


Fig. 19—Upper and lower confidence interval.

of the size. Both the asymmetry and the decrease in the size of the confidence interval are evident. Note that for the smaller areas the effect of the asymmetry is greater. Figure 20 is the same type of plot for an observed value of $30 of CPPAP for an area with 250 assigned pairs for differing numbers of months. Note that for this very small area, the confidence limits are quite wide and the effect of the asymmetry is much greater than that seen in Fig. 19.

Table III can be used to help decide an appropriate size for allocation areas and an appropriate length of time for data collection. For a given size and time, the confidence intervals for various observed values of CPPAP can be read from Table III. For example, if allocation areas are created of size 500 assigned pairs or larger, and if data are collected for 12 months or longer, then an observed CPPAP of $20 would give a confidence interval of $14.25 to $27.92—or a shorter interval if the area is

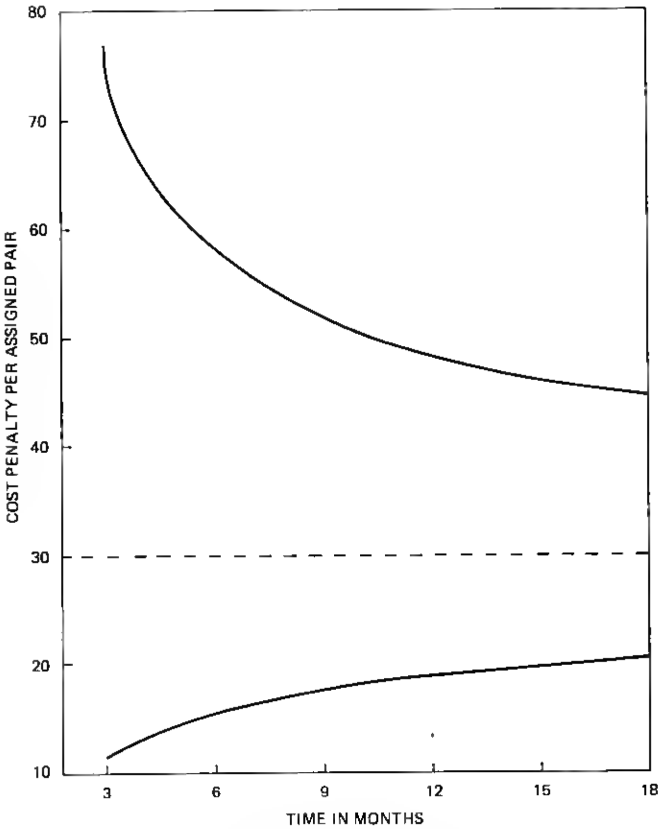


Fig. 20—Upper and lower confidence interval.

larger or the data collection period longer. If the uncertainty in the "true" CPPAP represented by this interval is acceptable, then allocation areas could be sized to a minimum of 500 pairs with data collection for a minimum of 12 months. The uncertainty resulting from alternative values of size and time can be checked in this way using Table III. When forming allocation areas in a district and determining the length of time for data collection, the minimum size and time should be chosen so as to produce results precise enough for the decision making needs of the district.

4.2. *Extending results to individual areas*

The basic results presented in Table III are given for only three values of the measured CPPAP, six different collection periods, and six area sizes. The first and most straightforward extension of this analysis to different areas and collection periods involves extending the tables using eq. (4) or by linear interpolation of the given table values. As can be seen from

Figs. 19 and 20, any linear interpolation is more valid for the range of the table associated with longer collection times and larger collection areas. This is simply because the effect of the transformation is more linear for this range of values.

In the event that users of CPPAP data feel that their areas are significantly different from the Prototype District, which is the basis of Table III, there are several ways in which this analysis can be modified. First, the constant associated with eq. (6) can be re-estimated using the techniques described in Section 4.1. While the estimation of the weights in the regression is somewhat more complicated than ordinary least squares, most commercially available statistical computation packages allow for this type of estimation. Having computed the constant which relates variability to size of area, it is a simple matter to generate tables analogous to Table III.

However, the logarithmic transformation of CPPAP used here for analysis might not always satisfy the desired assumptions. In this case a more exploratory analysis should be undertaken. Unfortunately, such an analysis will require additional statistical computation and display. Tbe sequence of steps discussed in Section III can serve as a guide for the analysis, and for checking the appropriateness of various transformations. Finally, it is possible that no appropriate transformation will be found. Then the method of analysis employed in this section will not be adequate.

V. ANALYSIS OF AFTER CONVERSION DATA

5.1. Description of analysis

A major concern in the conversion of serving areas to SAC is whether or not the projected savings are being realized. To help answer this question the cost penalty data in the periods after conversion are examined. A regression equation is developed which models the after conversion costs in terms of before and during conversion variables as well as the time since conversion. The most important result shows that the cost penalty continues to decline for the period immediately following conversion. The implication of these findings on conversion analysis is that to adequately assess the effect of conversion, cost data must be collected for a period of nine to twelve months after conversion.

One might assume, *a priori,* tbat there will be differences in the converted areas but that such differences would not be related to the before or during conversion periods. These areas were all rehabilitated using the same FAP guidelines, so they should start off on the same footing. Differences might be related to installer productivity or activity, or geographic considerations of the areas. However, data on such variables are outside the scope of the Prototype District Data Base and are not currently available. It is of interest to know to what extent after con-

version behavior might be explained, and the analyses of this section are directed at using variables available in the data base to this end.

Since the logarithmic transformation of the before conversion data satisfied straightforward assumptions needed for analysis (see Section III), one might expect this transformation also to be reasonable for the after conversion data unless there are some "structural" changes in the after conversion period. Our analyses do not indicate any such change, so the quantity analyzed here is $y = \ln(\text{CPPAP} + 1)$. Ten of the 23 allocation areas were converted, and each of these areas has from 1 to 20 months of after conversion data. The total number of values (areas · months) is 100.

We search for a linear description of the 100 y's of the form

$$y_{ij} = a_0 + a_1 x_{1ij} + a_2 x_{2ij} + \ldots + a_\ell x_{\ell ij} + e_{ij} \tag{7}$$

where i denotes area; j denotes month; x_1 is some descriptive or explanatory variable with value x_{1ij} for the ith area and jth month; similarly for $x_2, \ldots, x_\ell$; and e_{ij} is the residual which is unexplained, and which should not be related to any available variable. In accord with the analysis in Sections III and IV, we assume that $\text{var}(e_{ij}) = \sigma^2/p_i$. Thus, all regressions discussed here are weighted regressions with weights inversely proportional to the square roots of these variances. The problem is to find a good but parsimonious set of variables $x_1, x_2, \ldots, x_\ell$.

5.2. Fitted regression equation

Three classes of potential descriptive variables x are considered. First are variables which give some characteristics of areas, where these characteristics can be observed before the after conversion period. Such a variable has a fixed value for each area (i) across months (j). Examples include the distance of an area from the central office, the size of an area, and the average cost penalty for an area before conversion. The second class of variables concerns seasonal cycles across months. Such a variable has a value depending on the months (j) but is constant for each area (i). The third class consists of the single variable giving the number of months since conversion of that area; thus $x_{ij} = k$, where month j is k months past the conversion date of area i.

Consider the first class of variables. The most powerful such variables would be a set of 10, with each variable having some non-zero value in one area and the value zero in all other areas. This gives a one-way analysis of variance model, with the area corresponding to the treatment or groups. Doing this, one obtains an $R^2 = 0.28$. This means that 28 percent of the variation in the y's can be explained by differences between the areas.

The fit is improved substantially ($R^2 = 0.37$) by adding to this model the variable which measures the number of months since conversion.

However, the further addition of variables allowing different values for different months—the seasonal or time effect variables—improves the fit only negligibly. Thus, use of all the variables available here would result in a model describing about 40 percent of the variability in the after conversion values. Although this is not a large percentage on an absolute basis, it is also not negligible, especially considering that this is variability over months and areas after conversion to SAC.

Now we would like to go further and discover specific characteristics of the ten areas and specific variables that would give a simpler but still relatively good descriptive model. The following eight variables measuring characteristics of the areas were considered: the size of an area, as measured by the number of assigned pairs; distance to central office along feeder cable, measured in kilofeet; area mean before conversion; area standard deviation before conversion; area mean during conversion; area standard deviation during conversion; number of months of before data available; and number of months during conversion. The above one-way analysis of variance implies that the maximum descriptive power of any subset, or transformations, of these variables is 28 percent.

In order to find a small but good set of variables and transformations, extensive regression analyses were done, including stepwise calculations and C_p analysis.[10] As is often the case in such problems, no small set of variables clearly stands out as the unique "best" regression equation. Correlations between explanatory variables can permit several different sets of variables to fit the data approximately equally well. We will now discuss one simple model that does fit these data reasonably well.

Variables included in the model are the following: number of months since conversion; during conversion mean; during conversion standard deviation; and number of months before conversion. The fitted regression equation is summarized in Table IV, which gives the regression coefficients, the estimated standard errors, and the t-values for testing each coefficient equal to zero. The R^2 is 0.35 with residual standard error of 0.44, compared to a standard deviation of 0.54 for the dependent variable. Thus, use of only four variables gives a fit nearly as tight as can be obtained when using all possible explanatory variables available here.

Table IV — Fitted equation for after conversion data*

$y_{ij} =$	1.60	$-0.044x_{1ij}$	$+0.44x_{2i}$	$-1.13x_{3i}$	$-0.038x_{4i}$
Standard error	0.36	0.009	0.14	0.37	0.010
t-statistic	4.41	-4.89	3.19	-3.07	-3.80

* $y_{ij} = \ln(\text{CPPAP} + 1)$
x_{1ij} = number of months since conversion
x_{2i} = during conversion mean
x_{3i} = during conversion standard deviation
x_{4i} = number of months before conversion
(x_2, x_3, and x_4 are all the same over all months; hence, the time subscript j is omitted.)

No monthly time variable or cyclic time effect is included, since the analysis showed that they had no additional explanatory power.

Examination of various residual plots is important in determining the adequacy of this fit. Figure 21 gives a partial residual plot[11] for the number of months since cut-over (x_1) variable. The variable plotted on the vertical axis is the residual from the regression fit plus the contribution from this variable. Thus, one expects the points to scatter about a straight line with slope equal to the regression coefficient for x_1, here −0.044. This figure does not suggest any serious inadequacy in the fit as far as this variable is concerned. Partial residual plots and residual plots for the other variables, normal q-q plots, and various box plots of the data and residuals were also examined. They did not show anything particularly noteworthy.

Consider the interpretation of the variables in the fitted equation. For variable x_1, the number of months since cut-over, it is not surprising that the level declines over time after the conversion is completed, since unknown cable troubles and defective pairs will be discovered and corrected. Figure 21, introduced above, shows graphically that there is a steadily decreasing trend as the number of months since cut-over increases. There is not an instantaneous decline to a low, constant level.

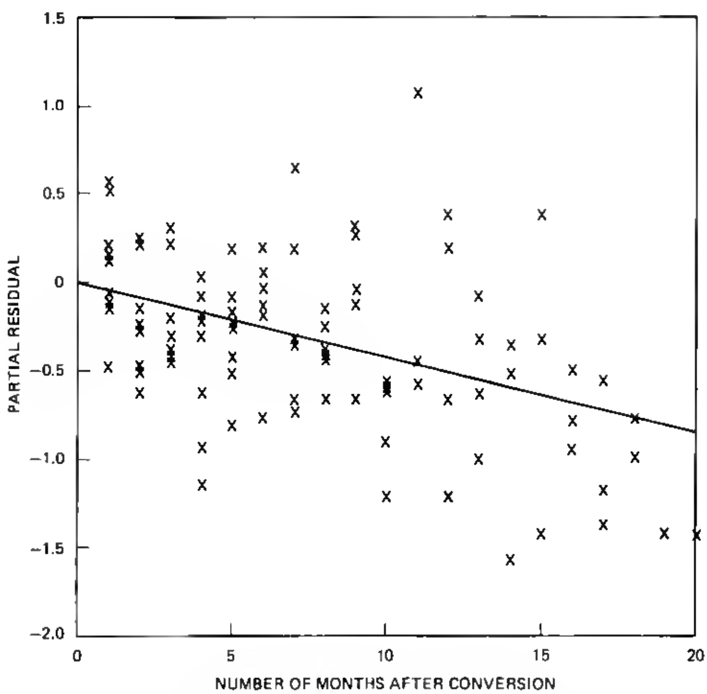


Fig. 21—Values of ln(CPPAP + 1) after conversion.

Moreover, this variable (x_1) appears with approximately the same negative coefficient in all "reasonably fitting" sets of variables, while other individual variables are not so strongly needed in order to obtain an adequate fit. For variable x_2, the during conversion mean, it seems reasonable that a higher during conversion period (a proxy for the complexity of the conversion activity) will be associated witb a larger after conversion level. However, the interpretations for the during conversion standard deviation (x_3) and the number of months before conversion (x_4) are not as straightforward. For example, one could speculate that areas with a high level of during variability have spots of local congestion causing occasional high costs (i.e., RE's LST's, WOL's, etc.). A large standard deviation implies that there are also months in which costs are low. It is just this type of area that can show large savings (and lower values of CPPAP) after conversion via FAP. The number of months before conversion could be a proxy for the ranking of the converted areas. Presumably, the worst areas would be converted earlier. Hence, the better areas are converted later and the post conversion costs of the better areas are lower (other things being equal).

VI. ACKNOWLEDGMENT

We are indebted to many members of Department 4511 for their time and effort in explaining concepts and issues pertaining to the loop plant. Special thanks are due Nancy Basford who never failed to respond helpfully to our many queries.

REFERENCES

1. G. W. Aughenbaugh and H. T. Stump, "The Facility Analysis Plan: New Methodology for Improving Loop Plant Operations," B.S.T.J., this issue.
2. J. O. Bergholm, private communication.
3. J. O. Bergholm and P. P. Koliss, "Serving Area Concept—A Plan for Now With a Look to the Future," Bell Laboratories Record, *50*, No. 7 (August 1972), pp. 212–216.
4. N. L. Long, "Loop Plant Modeling: Overview," B.S.T.J., this issue.
5. G. W. Aughenbaugh, N. L. Basford, D. M. Dunn, A. E. Gibson, and J. M. Landwehr, private communication.
6. A. E. Gibson, personal communication.
7. H. P. Friedman, et al., "A Graphical Way of Describing Changing Multivariate Patterns," Proc. of the Comp. Science and Statistics 6th Annual Symposium on the Interface (1972), pp. 56–59.
8. J. W. Tukey, *Exploratory Data Analysis*, Reading, Mass.: Addison-Wesley Publishing Co., 1977.
9. M. B. Wilk and R. Gnanadesikan, "Probability Plotting Methods for the Analysis of Data," Biometrika, *55* (1968), pp. 1–17.
10. C. L. Mallows, "Some Comments on C_p," Technometrics, *15* (November 1973), pp. 661–676.
11. W. A. Larsen and S. J. McCleary, "The Use of Partial Residual Plots in Regression Analysis," Technometrics, *14* (August 1972), pp. 781–790.